DOCUMENT RESUME

ED 310 058 SP 030 698

AUTHOR Porter, David B.
TITLE Course Critiques: What Students Can Tell Us about Educational Efficacy.
PUB DATE 88
NOTE 7p.; In: Proceedings of the Human Factors Society, 32nd Annual Meeting, 1988. p473-77.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Students; *Course Evaluation; *Evaluation Methods; Higher Education; Likert Scales; *Student Evaluation of Teacher Performance; *Teacher Effectiveness; Teaching Styles

ABSTRACT

A course critique based on a multi-channel model of education has been developed and administered to 1,599 students enrolled at the U.S. Air Force Academy in the 1988 spring semester. The single-page instrument asks students to evaluate contributions the overall course made to their personal enjoyment, critical thinking, and subject knowledge, as well as rate the contributions of 10 course elements such as instructor, text, and classmates. The utility of this approach is shown by data analyses at three levels of evaluation: department, course and instructor. Several pedagogical implications are discussed. (JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED310058

COURSE CRITIQUES: WHAT STUDENTS CAN TELL US ABOUT EDUCATIONAL EFFICACY

David B. Porter

USAF Academy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. B. Porter

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

SP 030 698

BEST COPY AVAILABLE

COURSE CRITIQUES; WHAT STUDENTS CAN TELL US ABOUT EDUCATIONAL EFFICACY[1]

David B. Porter
USAF Academy

A course critque based on a multi-channel model of education has been developed and administered to all students enrolled in behavioral science courses at the US Air Force Academy for the past three semesters. The utility of this approach is shown by data analyses at three levels: department, course and instructor. Several pedagogical implications are discussed.

Many educators mark the end of quarters or semesters by asking their students to rate their "teaching effectiveness." The use of such evaluations is pervasive but the current literature shows a disconcerting amount of confusion, argument and controversy. On the one hand, Phye (1984) assures us that "little doubt remains that well-constructed student evaluations can provide valid, reliable information..." (p.92). On the other, Sommers (1981) revealed that many of the actual instruments were described by other researchers as "confused," "hit or miss," "whimiscal" and as generally having little to do with teachers' actual performance. Nonetheless, there is general agreement and considerable evidence that class-average student ratings are reliable and stable as well as multidimensional (Marsh & Hocevar, 1984; Howard, Conway & Maxwell, 1985; and Gaski, 1987). However, there is also considerable controversy about exactly what the underlying constructs being measured are and how they relate to education. For example, Howard et al. (1985) employ convergence to validate their measure of "teaching effectiveness." In a well-supported review, Gaski (1987) uses the same data to argue that more appropriate labels for the general factor they discovered might be "naive erroneous perception of teaching effectiveness, visceral reaction or ease of grading" (p. 327).

What seem to be lacking (and are major contributors to subsequent controversies) are clear a priori statements of educational theory and explications of the logic which connects the ratings instructors receive to the overall educational process. Impressive and powerful statistical procedures may have made it all too convenient for us to get the data (and results) first and worry about the coherence of underlying educational theories later. The trepidation with which many faculty members view student critiques is well-founded. Despite the controversies, theoretical paradoxes and often equivocal results, the fact remains that "student ratings tend to be the only tangible source of instructional evaluation information in the majority of colleges and universities, both here and abroad" (Aleamoni, 1981, p.110). Decisions concerning curricula and faculty by both administrators and students are likely to be influenced by data from course critiques; we need to understand them better.

One way to approach the problem is to first define our terms. Education is the process of developing knowledge, skill and character through experience. Efficacy is simply the power to produce the intended results. The "educational efficacy" of any system clearly depends on achievement of its "intended results." Although most research on teaching effectiveness has relied on "student achievement" as the sole validation criterion (Cohen, 1982), most professional educators are uncomfortable with the idea that the purpose of education is merely for students to perform well on standard multiple-choice exams. It is useful to expand the criterion. Perhaps a more robust definition would be that the purpose of education is to provide our students with the knowledge, skills and motivation necessary for them to make significant intellectual contributions to society or our particular discipline as graduates. Although such a broad goal cannot be measured immediately, it provides a solid foundation and useful framework for understanding the measures we receive from our students. This framework is the Multi-channel Model of Education (Figure 1).

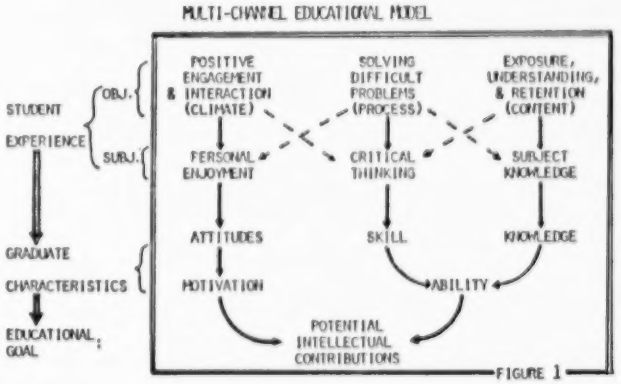

FIGURE 1

The model depicts a general temporal flow from top to bottom. Objective occurrences lead to subjective experience. Aggregates of these educational experiences influence the subsequent characteristics of our graduates. Ability depends on having stored needed information (knowledge) and also having

acquired the active processes (skill) to effectively combine new information with existing knowledge. Although ability and motivation do not guarantee substantial contributions (i.e., they are not sufficient), both of these constituents of performance are clearly necessary. Objective occurrences (the things that go on in our classrooms) can be viewed from three distinct perspectives: content (what), process (how) and climate (psychologically where). Although each aspect of objective experience might influence all three subjective outcomes the solid arrows depict the strongest and most direct relationships.

Several important points emerge from a comparison of this model and the more simple models implied by much of the current research. Many studies of student critiques assume that only the far right knowledge channel is educationally relevant. The multitrack model recognizes the necessity of each channel; relative importance is determined by the scarcity of "positive flow" in any of the three channels (i.e., the most important ingredient is the one that is most lacking). One consequence of our failure to recognize the viability of all three channels are educational systems which each year induct freshmen high in motivation but low in ability and graduate seniors high in ability and low in motivation. From a broad perspective, the net contribution of such systems is zero.

Each of the three channels require different types of measurement. Traditional, multiple-choice tests provide a viable measure of students' retained knowledge but may tell us little about their ability to apply their knowledge to real problems and even less about their intrinsic motivation (i.e., willingness to use these skills). Problem-solving exercises or large projects are more appropriate indicators of skill than multiple-choice tests. Student behaviors such as the choice of challenging goals and high persistence in the face of adversity are much better indicants of "adaptive motivational patterns" (Dweck, 1986). From this perspective "the validity" of student critiques depends on the particular questions the critique contains as well as the measures we choose for validation. For example, if the critique focuses on students' affective reaction to the psychological climate of the classroom, it may show only a weak relation with students' subsequent performance on objective achievement tests. On the other hand, we might expect stronger support if our criterion was the percentage of students choosing to take more courses in the discipline (or with the same instructor).

Now that we've partitioned the hypothetical causal relations leading to our educational goal into separate channels, it is time to disabuse ourselves of another assumption implied by some other studies of "the validity" of student critiques. The teacher is not the only influence on student learning. As Aleamoni (1981) reminds us, many "course elements such as the instructor, textbook, homework, course content, (and) method of instruction... all serve to change student behavior" (p. 111). McKeachie, et al. (1981) points out "the most effective method of teaching depends on the goal, the student, the content, and the teacher." (p.63). From this perspective, an interesting educational implication emerges. More is not always better and may even be worse. For example, some instructors outline each assigned reading and then provide their students with an organized and entertaining inclass presentation. Such teaching behavior is likely to result in both acknowledgment from students (high critique ratings) and perhaps improved performance on some measures of achievement. However, from a broader perspective, it might be argued that by denying students the opportunity to develop necessary study skills and the satisfaction of undertaking the more difficult and challenging academic task of structuring the material themselves, these teachers are not the pedagogical heroes they appear to be from their critiques and students' performance. Although chronic poor performance or low student ratings often suggest problems with instruction, the assumption that good performance (on a particular test) and good ratings from students are sufficient to define good instruction is clearly faulty.

Many authors stress the necessity of specifying the intended use of critiques prior to their development and administration (Bowers, 1980; Sommer, 1981; Aleamoni, 1981). One advantage of presenting a general educational model, is that it allows us to use a single instrument but complementary analyses to interpret the data from different perspectives and at different levels. Three such levels of analysis will be presented here.

Regularities in the pattern of student responses across all sections and courses provide insight into the general educational process. The general multi-channel model presented earlier suggests several specific relationships which should emerge from the data. The differential responses of students in various courses are also important. At this level of analysis, the focus is on the curriculum and courses rather than on the instructor per se. This type of analysis is particularly relevant to the evaluation of experimentation with alternative educational approaches. A third level of analysis relates to the individual instructor. Many existing critiques offer several subscales which are considered to add to the teachers perceived effectiveness as well as a single global rating. The instrument employed in this study could be used in this way; however, it also

allows us to explore an intriguing alternative.

In a recent article, Hudak and Anderson (1984) studied the relationship between teaching style and several measures of teaching effectiveness. Drawing on Broudy's pedagogical theory, they introduce three distinct teaching styles: philetic, heuristic and didactic. Philetics emphasize their relationship with students. These teachers are warm, approachable and value students' views. In contrast, heuristic instructors emphasize ideas and stress conceptual integration with the goal of teaching students to think rather than merely memorize. The third style is "didactic" and is the most traditional. This style emphasizes Memorization, drill and detailed recitation knowledge often characterize this approach. These three styles are considered to be relatively distinct and have both advantages and disadvantages. The novel thesis implied by this approach is that there may not be "one best style of teaching." The best faculties may be those that obtain diversity by encouraging instructors to develop a style which is congruent with their personal values, pedagogical skills and the particular course they teach.

Method

Students' critiques have been in use by instructors in the Department of Behavioral Sciences and Leadership at the USAF Academy for many years. Several years ago, the decision was made to develop a single instrument for use by all instructors and in all courses. Since instructor "acceptance" was deemed to be at least as important as "quality," maximal participation from the faculty was encouraged. Through several iterations, a single-sheet critique was developed. Students are asked to rate each of 24 questions relating to their instructor's behavior on a 5-point Likert scale. There are also seven questions which reflect several potential criteria and moderator variables (e.g., hard work, class participation, perception of fairness and midterm grade). In the final portion of the critique, students are asked to rate the contribution of each of 10 course elements as well as the overall course on each of three outcomes: personal enjoyment, ability to think critically and knowledge about the course's subject. Students also used a 5-point scale to complete this 11 by 3 outcomes matrix.

Critiques were administered to 1599 students during the last lesson of the 1988 Spring semester. In addition to departmental averages and standard deviations, responses were aggregated for each of 15 courses and 37 instructors. Overall correlations within and between critique and matrix items were computed.

Results

Department Averages

The large sample size lends great stability to the pattern of responses. In fact when the results from the Spring 1988 critiques were compared with those obtained during Fall 1987, reliability exceeds .98. The three most general items are the matrix responses which ask students to evaluate the contribution the overall course made to their personal enjoyment, critical thinking and subject knowledge. The average ratings for both enjoyment and critical thinking were 2.66 on a scale ranging from 0 to 4 (i.e., 2.0 would be neutral). Students ratings of courses contributions to their knowledge (2.86) was significantly (but not substantially) higher ($z = 7.99$, $p < .01$). Of greater theorectical interest are the average contribution profiles of several different pedagogical elements common to all our courses. These are shown in Figure 2.

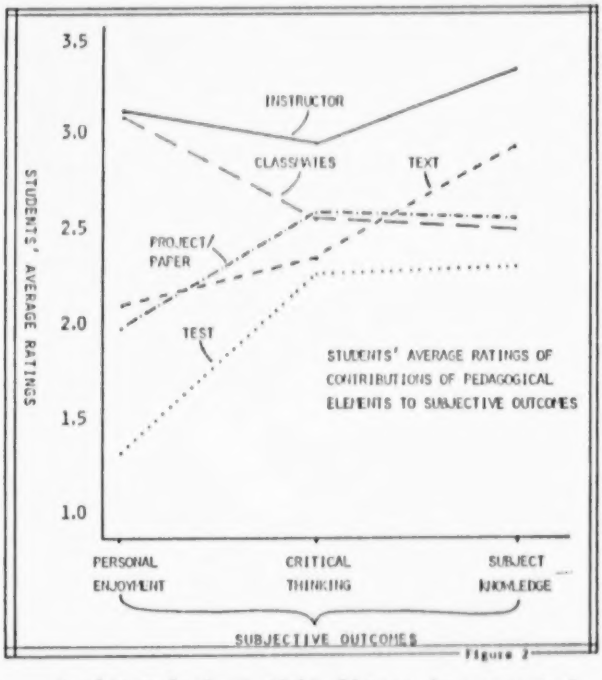

Figure 2

A closer look at this figure is warranted. Students rate the instructor and their classmates as the two strongest and nearly equal contributors to their personal enjoyment in the classroom. Texts and projects are rated as having a neutral effect and tests are seen to have a negative influence on this outcome. There is an interesting difference in the respective contribution profiles for the text and student projects. While texts are seen to contribute more to knowledge, projects make a greater contribution to critical thinking. The graph also suggests students perceive instructors as making the greatest contribution to both critical thinking and course knowledge.

Course Averages

Course averages provide useful information to our senior faculty in making curricular decisions. The data clearly show the strong positive response of students to most of our courses. Not only do course average data indicate the few courses which need attention, they provide valuable diagnostic information about which aspect of the course may be causing problems (i.e., the text may be seen to contribute little to student knowledge, discussions may not be very enjoyable or the course project did not contribute to critical thinking).

These data are also valuable in evaluating alternative approaches to education. One example will illustrate this point. Inter-Disciplinary Education at the Academy (IDEA) is an innovative program initiated several years ago to enhance educational integration and to explore the advantages of encouraging collaborative learning among students. Students in a particular class in one course (e.g., Physics) would all be assigned to the same section in a different course (e.g., English or Psychology). Instructors were encouraged to attend the alternate class with their students and discuss ways in which the two disciplines were related. Many of these pairings also sought to enhance group cohesion by assigning integrative group projects and establishing interdependent grading systems which rewarded group as well as individual performance.

This concept was applied last semester to two courses within our department. One section of 22 students remained together (and had the same instructor) for a course in leadership and one in advanced research design and statistics. These students were randomly selected from all students who were enrolled in both courses for the Spring term. Their initial overall grade point average did not differ significantly from other students enrolled in either course. Performance on common graded evaluations of this group was considerably better than that of other students in either course ($t(21) = 2.45$, $p<.025$ for the leadership course and $t(21) = 4.06$, $p<.01$ for statistics). However, many questions remained. IDEA students' ratings of both courses were not significantly higher than the ratings from students enrolled in other sections. Contribution profiles for the instructor, text and classmates, however, did show distinct differences from other sections. The texts in both courses were seen to be much more enjoyable ($t(21) = 4.30$, $p<.01$) but not as contributing any more to students' critical thinking or their knowledge of the subject. IDEA students did not rate their classmates as being significantly more enjoyable than non-IDEA students rated their classmates, However,IDEA students did rate their classmates as contributing much more to their critical thinking ability ($t(21) = 3.63$, $p<.01$) and their knowledge of the subject ($t(21) = 2.94$, $p<.01$). Since grades were primarily determined by students' performance on standard multiple-choice examinations which tested subject knowledge, it appears that students attributed their improved performance primarily to their classmates. This evidence provides strong support for the hypothesis that under appropriate conditions students can significantly enhance their own educational experience.

Instructor Averages

One of the primary purposes for student critiques is to provide instructors the information they need to evaluate and improve their own teaching. Toward this end, each instructor is given data from each of their sections as well as their average ratings across all sections. For comparison, they are also provided current averages for the course they taught and overall department averages. Instructors are given some written guidance concerning interpretation but are also encouraged to discuss results with colleagues and supervisors.

Work with the teaching styles discussed earlier is still in its initial stages but has already yielded some intriguing results. Students' ratings of instructors contributions to enjoyment, critical thinking and knowledge are nearly isomorphic with Broudy's pedagogical styles. Didactic instructors should be seen to contribute the most to knowledge; heuristic to critical thinking and philetic to personal enjoyment. Students' ratings of these three contributions are strongly interrelated (all three correlation coefficients are between .62 and .65). Although it appears that about half the variance in each measure is explained by either or both of the other two, this implies that half of the variance is unique. A cursory review of the overall correlation matrix provides some interesting convergent evidence. Some instructor behaviors are strongly related to all three outcomes (e.g., "effective communication," and "interesting classroom presentations" correlate with all three outcomes with r's between .50 and .65). Other behaviors appear to be associated with particular outcomes (e.g., "using humor" was closely associated with enjoyment ($r=.60$); asking provocative questions with critical thinking ($r=.54$); and knowing the material with subject knowledge ($r=.56$).

By taking each instructors' most highly rated contribution (i.e., to enjoyment, thinking or knowledge) and subtracting the average contribution rating of the other two outcomes, instructors could be segregated into three stylistic categories. Of the 37 faculty members, 31 were didactic, 5 were philetic and 1 heuristic. There was no relationship between teaching style and student performance, however, teaching style showed a strong relationship

with two other factors. Didacticism was negatively related to seniority (r=-.548, p<.01) and even more strongly related to students expressed desire to take another course from the instructor (r=-.645, p<.01). It appears that with experience instructors become less didactic and as a result make more positive impressions on their students. These results replicate Hudak and Anderson's (1984) findings.

Discussion

Students' critiques of their instructors are both ubiquitous and controversial. The absence of explicit educational models supporting particular critiques contributes to both the ubiquity and controversy of their administration and interpretation. The multi-channel educational model provides an integraded conceptual framework to support the use of student critiques in the evaluation of educational efficacy. This model assumes that what goes on in the classroom influences students' subjective reactions which, in turn, aggregate to affect the characteristics of graduates. Separate channels for attitudes, skills and knowledge are hypothesized. An outcome contribution matrix based on this model was developed and has been administered to students enrolled in behavioral science courses at the Air Force Academy for the last three semesters.

Data from these critiques examined at three levels to demonstrate the potential utility of this approach. Departmental averages showed distinctive profiles for different pedagogical elements (i.e., the instructor, classmates, tests, etc.) General logical consistencies of these results suggest students were able to conceptually separate the three subjective outcomes as well as discriminate the various course elements. A valuable use of these data at the course average level was also shown. In particular, an alternative educational approach involving collaborative learning received convergent support from the data this instrument provided. The outcomes matrix also offers the opportunity to provide "stylistic" feedback to instructors. Results from this analysis suggest younger instructors tend to be more "didactic" than their more senior colleagues. Although didacticism was unrelated to students' performance, it was strongly (and negatively) related to student satisfaction.

A great deal of validation research with this model and the matrix critique used to instantiate it remains. The purpose of this brief paper was to introduce the general approach and the multichannel educational model and to demonstrate some potential uses.

Note

1. The views expressed in this paper are those of the author and do not necessarily reflect those of the USAF Academy, US Air Force or Department of Defense.

References

Aleamoni, L. M. (1981). Student ratings of instruction in National Council on Measurement in Education, Handbook of teacher evaluation. Beverly Hills, CA: SAGE.

Bowers, J. K. (1980, April). A longitudinal study in academic evaluation. Teaching of Psychology, 99-102.

Cohen, P. A. (1982, April). Validity of student ratings in psychology courses: A research synthesis. Teaching of Psychology, 9, 78-82.

Dweck, C. S. (1986, October). Motivational processes affecting learning. American Psychologist, 41, 1040-1048.

Gaski, J. F. (1987). On construct validity of measures of college teaching effectiveness. Journal of Educational Psychology, 79, 326-330.

Howard, G. S., Conway, C. G., & Maxwell, S. E. (1985). Construct validity of measures of college teaching effectiveness. Journal of Educational Psychology, 77, 187-196.

Hudak, M. A. & Anderson, D. E. (1984, October). Teaching style and student ratings. Teaching of Psychology, 11, 177-178.

Marsh, H. W. & Hocevar, D. (1984, Summer). The factorial invariance of student evaluations of college teaching. American Educational Research Journal, 21, 341-336.

McKeachie, W. J., Pintrich, P. R., Lin Y. G., & Smith, D.A.F. (1986). Teaching and learning in the college classroom: A review of the research literature. Ann Arbor, MI: Univ of Michigan (National Center for Research to Improve Post-Secondary Teaching and Learning.

Phye, G. D. (1984, April). Student performance and the evaluation of teaching effectiveness. Teaching of Psychology, 11, 92-95.

Sommer, R. (1981, December). Twenty years of teaching evaluations: one instructor's experience. Teaching of Psychology, 8, 223-226.